Environment Canada
Environnement Canada

BENTHIC INVERTEBRATE COMMUNITY EXPERT WORKING GROUP FINAL REPORT

RECOMMENDATIONS FROM CYCLE 1 REVIEW

PREPARED FOR THE
EEM TECHNICAL MANAGEMENT COMMITTEE

EEM/1997/7

Canada

Publications relating to Environmental Effects Monitoring (EEM):

Aquatic Environmental Effects Monitoring Requirements (revised EPS 1/RM/18). Environment Canada. 1997. EEM/1997/1

Pulp and Paper Aquatic Environmental Effects Monitoring Requirements. Environment Canada. 1997. Annex 1 to EEM/1997/1

Tracer Expert Working Group Report. EEM Technical Management Committee. 1997. EEM/1997/2 is also published as: Ali, N et al, 1997. Exposed or not exposed: the effectiveness of using tracers to verify exposure of fish to pulp and paper mill effluent. Canadian Technical Report of Fisheries and Aquatic Sciences 2162.

Dioxin Expert Working Group Report. EEM Technical Management Committee. 1997. EEM/1997/3

Usability Expert Working Group Report. EEM Technical Management Committee. 1997. EEM/1997/4

Toxicology Expert Working Group Report. EEM Technical Management Committee. 1997. EEM/1997/5

Fish Survey Expert Working Group Report. EEM Technical Management Committee. 1997. EEM/1997/6

Benthic Community Expert Working Group Report. EEM Technical Management Committee. 1997. EEM/1997/7

Pulp and Paper Technical Guidance for Aquatic Environmental Effects Monitoring. 1997. EEM/1997/8

Copies are available at:

National EEM Office
Science Policy and Environmental Quality Branch
Environmental Conservation Service
Environment Canada
Ottawa, Ontario, Canada
K1A 0H3
Tel.: (819) 997-1535
Fax: (819) 953-0461
Email: eem@ec.gc.ca
Website: http://www.ec.gc.ca/eem/english/default.htm

103-05634/

Note

The contents of this report are the views of the members of the Environmental Effects Monitoring (EEM) Expert Working Group. As well, the recommendations in this report are the result of the Expert Working Group process and represent the conclusions of the Expert Working Group.

The recommendations and decision trees put forward in this report will be submitted to the EEM Technical Management Committee (TMC). This information will be used by the TMC to produce the TMC recommendations for updates to the regulatory requirements of the EEM program.

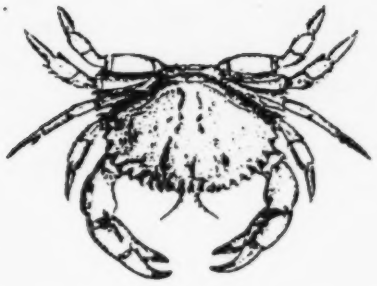
The documents which will contain the revised regulatory requirements will be called:

- Aquatic Environmental Effects Monitoring Requirements, EEM/1997/1
- Pulp and Paper Aquatic Environmental Effects Monitoring Requirements, ANNEX 1 to EEM/1997/1

Acknowledgments

The EEM Technical Management Committee (TMC) extends its thanks to the members of the Expert Working Groups for their contribution to the EEM program. The members of the group have produced their reports in a very tight time frame and they have displayed a willingness to work together to meet common objectives.

The review performed by the Expert Working Groups and their recommendations for enhancements to the program are essential ensuring that the EEM program remains based on sound science.



Benthic Invertebrate Community Expert Working Group

Technical Report on Pulp and Paper Environmental Effects Monitoring (EEM) Program

Review of Cycle 1 and Recommendations for Cycle 2

S. Megraw (co-chair) - Domtar Inc.
T. Reynoldson (co-chair) - Environment Canada
R. Bailey - U. of Western Ontario
B. Burd - UBC/Institute of Ocean Sciences
L. Corkum - U. of Windsor
J. Culp - Environment Canada
C. Langlois- Environment Canada
E. Porter - Environment Canada
D. Rosenberg - Fisheries and Oceans
D. Wildish - Fisheries and Oceans
F. Wrona - Environment Canada

August 7, 1997

Table of Contents

1.0 Executive Summary
2.0 Introduction
- 2.1 Background
- 2.2 Terms of reference
- 2.3 Activity and progress

3.0 Review of Cycle 1
- 3.1 Background
- 3.2 Overview of Cycle 1 results
- 3.3 Confounding Effects
  - 3.3.1 Historical impacts
  - 3.3.2 Multiple discharges
  - 3.3.3 Reference sites
  - 3.3.4 Sampling methods
- 3.4 Data Analysis
- 3.5 Summary

4.0 Recommendations for Cycle 2
- 4.1 Site Selection
- 4.2 Sampling methods
- 4.3 Data Interpretation
  - 4.3.1 Supporting field measurements
  - 4.3.2 Data measurement variables and analysis
  - 4.3.3 Data handling and screening
  - 4.3.4 Data manipulation
  - 4.3.5 Effect size
- 4.4 Technical Workshop

5.0 Available Alternative Methodologies and Approaches
6.0 Decision Tree
7.0 Future work and Research Needs
- **7.1 Technical Guidance Document**
- **7.2 Research**
  - 7.2.1 Mesocosm approach - field and laboratory
  - 7.2.2 Alternate sampling program designs using predictive modeling techniques (e.g., Reference Condition Approach)
  - 7.2.3 Application of the TRIAD approach
  - 7.2.4 Microbial characterisation
  - 7.2.5 Taxonomic resolution

8.0 References

1.0 Executive Summary

This report examines the results of the Cycle 1 benthic invertebrate community survey that was carried out as part of the Environmental Effects Monitoring (EEM) program for the pulp and paper industry. The Benthic Community Expert Working Group (BCEWG) based its review on the proceedings of various government/industry regional workshops, summary spreadsheets prepared by Environment Canada, the Cycle 1 experience of the BCEWG members, and summary reports prepared for the Atlantic, Quebec and Ontario regions. Major issues identified for review were, the ability to circumstantially relate invertebrate community response to mill discharges, the standardization of sampling methods, and the suitability of methods for data analysis.

Of the mills reviewed (112), 69% of the studies showed differences in species richness and/or abundance between reference and exposure sites, although fewer (28%) studies suggest that the differences were effluent-related. The interpretation of the remainder was confounded by factors such as multiple discharges and poor reference site selection.

A variety of sampling methods was used in Cycle 1 including different sample collectors, mesh sizes, and subsampling techniques. This lack of standardization makes it difficult to carry out a national or regional assessment of the environmental performance of mills.

The data analysis requirements were overly complex and restrictive in Cycle 1. The use of multivariate analyses was inappropriate for the study design and, in many cases, consultants derived their conclusions based on simple metrics rather than the results of the statistical analyses.

In addressing recommendations for Cycle 2, the BCEWG considered the first cycle issues being related to problems with a) study design and site selection, b) sampling methods, or c) interpretation. Within 'interpretation', several subcategories were identified including supporting field measurements, data measurement variables and analysis, data handling and screening, and effect size. With respect to effect size, no agreement was reached on what constitutes an "effect". Consequently, the BCEWG has begun to investigate developing effects guidelines for each of the measurement variables (community descriptors) proposed in this report.

In order to finalise the recommendations for Cycle 2, it is suggested that the contents of this report be reviewed by the consultants that participated in Cycle 1 and that a joint BCEWG/consultant workshop be held to ensure that the issues/problems have been adequately addressed.

Finally, this document has developed and tested a decision tree that will assist in planning studies for Cycle 2. It is not the intention of this report to replace the Technical Guidance Document; however, it is recommended that the Technical Guidance Document be reviewed for consistency with the recommendations of this report and where necessary modifications be made to it.

2.0 Introduction

2.1 Background

The pulp and paper sector is regulated with respect to waste disposal under the Pulp and Paper Effluent Regulations of the Fisheries Act, which prescribe effluent limits for biochemical oxygen demand (BOD), total suspended solids, and acute lethality to rainbow trout. Because of the possible inadequacy of the regulations in protecting fish and fish habitat due to the recognition that receiving environment responses are site specific (Wildish, 1983), a programme was developed to assess the effectiveness of the current regulations in protecting the receiving environment. The Environmental Effects Monitoring (EEM) Programme was initiated at locations in Canada where effluent is discharged to the aquatic receiving environment from either a pulp and paper mill or an off-site treatment facility that receives mill effluent. This programme was envisaged as occurring in a series of cycles. The first cycle, which was to be completed by April 1996, had the objectives of providing a baseline against which future cycles could be compared and providing a preliminary assessment of whether effects were evident in the receiving environment. The EEM programme was subdivided into a number of distinct elements: an adult fish survey, an invertebrate community survey, toxicity testing, tainting studies, detection of dioxins and the use of tracers. This report reviews the results of the invertebrate community component of EEM Cycle 1 and provides commentary and recommendations for Cycle 2.

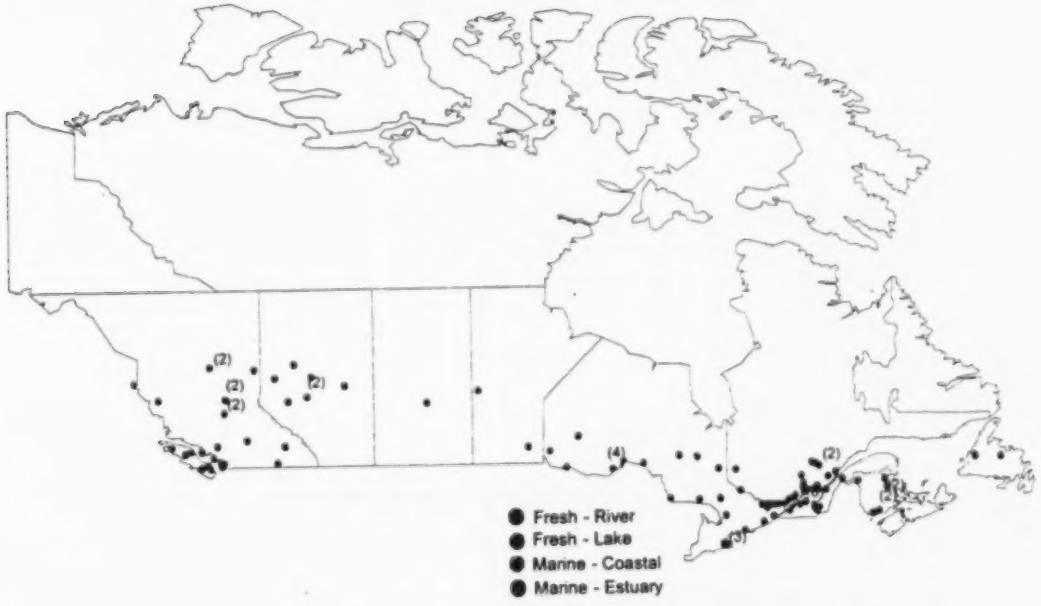


Figure 2.1. Location of pulp and paper mills in Canada participating in Cycle 1.

The overall objective of the EEM programme is to determine the spatial extent and the magnitude of the impact, if any, of pulp and paper mill wastes in freshwater, marine and estuarine environments. This objective was to be achieved by following the methods outlined in Environment Canada and Dept. of Fisheries and Oceans (1992). Overall the methods chosen were intended to be: 1. scientifically defensible, 2. cost effective, 3. flexible, 4. built on current research, 5. efficient of time and resources, 6. able to generate interpretable results, 7. able to provide a weight of evidence approach, and 8. able to give decision points based on comparative magnitudes of impact.

A total of 134 mills were included in Cycle 1 (Figure 2.1). The mills were sampled over the period September 1993 to November 1995. These data have been summarised and used in this report to review the effectiveness of the first cycle of EEM. In addition, regional mill summaries and the proceedings from a number of regional workshops were used in preparing this report.

2.2 Working Group Terms of Reference

The Benthic Community Expert Working Group (BCEWG) was formed in November 1996 and charged with the task of reviewing the results of Cycle 1 and providing recommendations for Cycle 2. Its principal responsibilities were:

- to review the concept and purpose of the various benthic community requirements under the EEM program,
- to identify difficulties encountered by the consultants in meeting the program requirements,
- to indicate situations where the requirements have been successfully met by the consultants,
- to account for situations where the national monitoring approach for Cycle 1 has proven to be unsuccessful and provide alternative approaches that would address site-specific difficulties or areas requiring additional research, and
- to provide a report identifying clear recommendations for Cycle 2 requirements.

The work of the BCEWG was guided by the following principles:

1. The basis of its recommendations were to be driven by regional evaluations of Cycle 1.
2. It respected a nationally consistent approach while allowing for flexibility in addressing site-specific issues.

3. Its recommendations followed the EEM program monitoring criteria and guiding principles as outlined in the General Program Requirements in document EPS 1/RM/18.

This report represents the initial attempt of the BCEWG to meet its responsibilities.

2.3 Activity and Overall Progress of the BCEWG

The BCEWG met on three occasions (November 5, 1996 - Burlington; January 31, 1997 - Montreal; Winnipeg June 12-13th 1997). The difficulties encountered by the consultants in meeting the program requirements were discussed, and the relationship between regional concerns and the proposed workplan provided a basis for recommending alternative approaches. The workplan developed by the BCEWG to address the issues arising from the findings of Cycle 1 included the following elements:

1. a reconsideration of the best way to report data. This required a database framework to ensure standard information collection among regions.

2. a modification and clarification of sampling design and statistical analyses to better meet the study objectives. This included a modification of the multivariate methods required, and recommendations for the use of selected biotic indices and univariate (ANOVA-type) tests.

3. a standardisation of the use of sampling gear within habitat types and a protocol for choosing sample sites.

4. an investigation of the use of a reference condition approach in freshwater environments and a gradient approach in marine/estuarine environments.

3.0 Review of Cycle 1

3.1 Background

The basic objective of the EEM programme is to detect the degree and spatial extent of changes in the receiving environment resulting from mill effluents. This demands that a prior decision be made as to what constitutes an important change in the receiving environment. However, such decisions incorporate input from both scientific and societal perspectives. Ecological science can determine the spatial extent of community change that a receiving environment may sustain. Ecological science has no opinion as to whether it is important to humans, and structural studies of this type provide no theoretical basis on which to judge its importance in ecosystem functioning. However, there is an inherent degree of uncertainty because of the normal variation associated with natural systems and our limited knowledge of how such ecosystems work. Therefore, within the cost and effort constraints of most assessment studies it is necessary to identify what constitutes an ecologically important change in an ecosystem. The societal component incorporates the degree of certainty that is required of the decision and the amount of divergence from the natural state that is to be accepted. Decision-making criteria that indicate such changes can use an array of attributes of the community including number of taxa, different metrics, overall structure and habitat descriptors.

Following from this, an EEM programme requires that the study design is such that an appropriate analysis of the data collected can measure the nature and magnitude of change in the ecosystem.

In assessing Cycle 1 of the EEM programme, it was first necessary to determine whether a change was detected. If a change was detected at a given mill, the evidence that the change was caused by mill discharges was assessed. It was then necessary to summarise the proportion of operating mills that are affecting the receiving system. Finally, to address the purpose of the EEM programme, the assessment must define whether current regulations are being effective in protecting receiving systems. This also requires criteria to be established for defining what is considered effective protection.

Therefore, in assessing the success of Cycle 1 it was necessary to identify:

1. Was a change in the receiving environment detectable?
2. How severe was that change in both spatial extent and deviation from natural conditions?
3. Could that change be circumstantially linked to the mill activities or were other factors confounding the interpretation?
4. Are the current regulations adequate?

The last issue on adequacy of current regulations was not addressed in this report.

3.2 Overview of Cycle 1 Results

In the first cycle, 134 mills were sampled to examine the benthic invertebrate communities downstream of pulp and paper mills. Of these results have been examined from 92 mills discharging to freshwater receiving systems and 29 to marine or estuarine waters. The major receiving systems included the Pacific Coast (11 mills), the Columbia system (2 mills), the Fraser River (7 mills), the Peace-Athabasca system (10 mills), the Churchill-Nelson system (3 mills), the Great Lakes (8 mills), the St Lawrence (14 mills) and the Atlantic Coast (14 mills). The remaining mills discharged to other smaller systems.

In reviewing the results of Cycle 1, the BCEWG relied largely on the following sources of information:

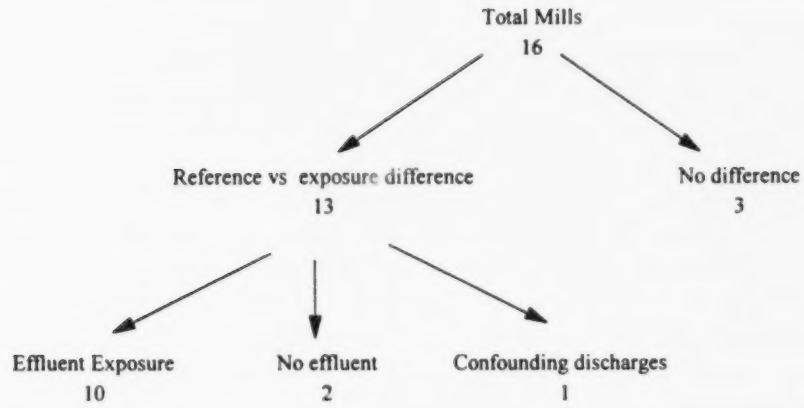
- Pacific/Yukon Region Institute of Oceans Science Workshop (May 27-30, 1996)
- BC Council of Forest Industries Workshop (June 20, 1996)
- Alberta Forest Products Association summary report (January 1997)
- Ontario Regional Government/Industry Workshop (September 10-11, 1996)
- Environment Canada/CPPA Integration Workshop (December 13, 1996)
- EC/CPPA Regional Workshop (February 4, 1997)
- Quebec regional workshop (March 20, 1997)
- summary spreadsheets provided by Environment Canada
- Cycle 1 experience of BCEWG members

Although useful in identifying the major issues and problems, there was, until recently, a general lack of sense as to how successful (or unsuccessful) Cycle 1 was in detecting changes. Part of the reason was that the information available to the BCEWG dealt with generalities and lacked specifics with respect to individual mill results. This is best illustrated by two regional summaries that were only released late in the Cycle 1 review process (Hamel et Vaillancourt, 1997; Parker and Smith, 1997). To-date (June 1997), the BCEWG have not received similar summaries from the other regions.

In summarising the results from the Atlantic region, Parker and Smith (1997) used species abundance and richness to show that some mills found no differences between reference and exposure sites, whereas others did. Of the 10 mills showing evidence of mill-related differences, the responses included increased abundance at 3, decreased abundance/richness at 6 and the presence of pollution tolerant organisms at 1 mill. At the time of the Cycle 1 study, two mills used "exposure" areas at the location of future outfalls so there was no effluent exposure. In both of these studies, decreased abundance/richness was observed, presumably a result of normal variation and suggesting that the reference sites were impoverished by environmental factors unconnected with

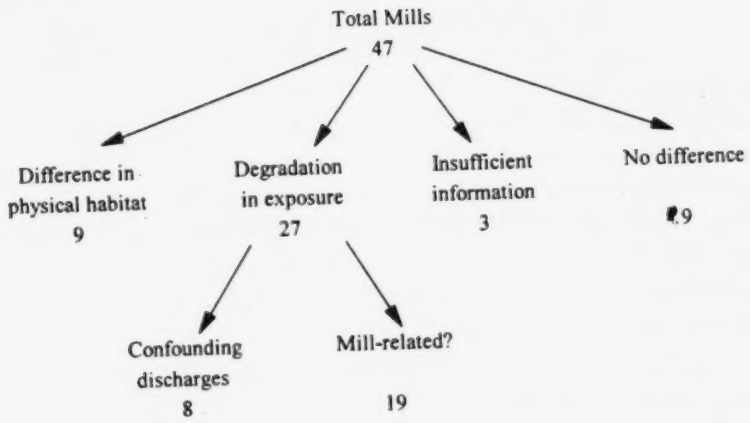
pulp mill effluents. Designation of near-field and far-field sites was not always stated so no assessment of spatial extent was attempted.

Atlantic Region



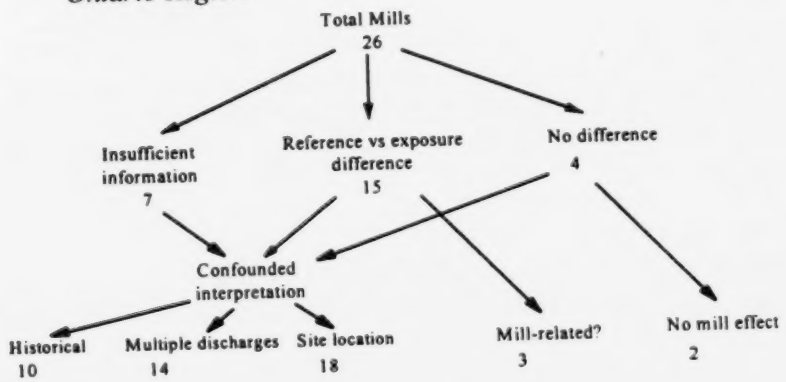
The results for the Quebec region are summarised below (Hamel et Vaillancourt, 1997). The criterion for "degradation" was not specified but, since the report indicated that only 11 of the 27 mills showed statistical differences, more than half the mills most likely relied on richness and/or abundance. Moreover, the types of responses in abundance/richness were not reported.

Quebec Region



Twenty-three individual review reports were available covering 26 mills in Ontario. These reports showed that, while the majority of studies (58%) identified differences between the reference and exposure sites, only in a few cases (3 mills) could these be circumstantially related to mill effluent. Most studies (21 of 26) differentiating between reference and exposure sites were unable to establish whether there were differences because of one or more factors, including multiple discharges, historical effects, or problems with site location.

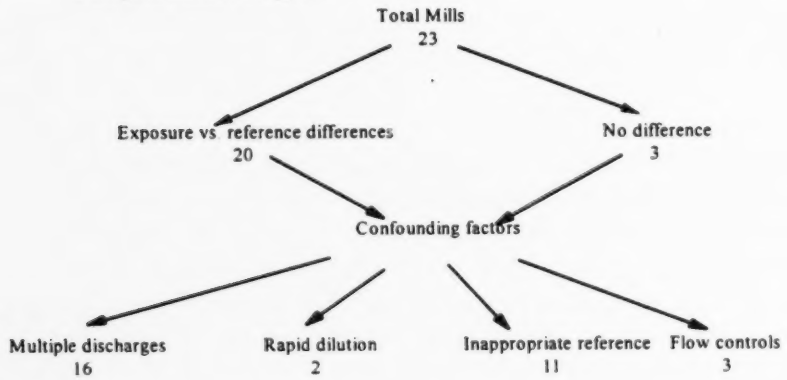
Ontario Region



In the Pacific region, 23 mills were reviewed. Two programs combined 2 mills and one program combined 4 mills for a total of 18 individual reports. Thirteen (13) mills discharge to marine and estuarine environments; 3 of these discharge to rivers where tidal reversal occurs at or just downstream of the mill diffuser. The remainder (10 mills) discharge to freshwater lakes or rivers. Subtidal depositional habitats were most commonly assessed in the marine and estuarine environments (11 mills); most freshwater mills assessed both erosional and depositional habitats in rivers and depositional in lakes and reservoirs. Sixteen mills (of 23 mills) assessed more than one

type of habitat. Methods for judging differences varied. In addition to the required total abundance and number of taxa, data were also compared using diversity indices (dominance, equitability, richness, Shannon-Weaver, Keefe-Bergersen), functional taxonomic groups, pollution- sensitivity, facultative/tolerant taxa, cluster analyses (Czekanowski, Bray-Curtis, Moriseta-Horn similarity coefficient), correspondence analysis, principle components analysis, discriminant analysis, or canonical correlations. Hypothesis testing was done using applicable ANOVA, ANCOVA, MANOVA, MANCOVA and bootstrapping. Confounding factors included appropriateness of reference sites (natural differences), multiple contaminant sources, hydroelectric dams, and historical deposits. Use of inappropriate reference areas was a particular problem in the marine environments. These problems made it difficult to judge whether differences observed in invertebrate communities were mill-related. About 8 mills clearly showed changes in species richness and 12 showed changes in abundance (5 elevated, 7 decreased). At least 18 mills showed signs of effluent exposure due to the presence of mill-related chemicals in the sediments.

Pacific/Yukon Region



From this simple analysis, the following pattern emerges for 84% of the mills participating in Cycle 1.

	No. of mills	Percentage (%)
Total no. of mills	112	
No differences	19	17
Incomplete information	10	9
Confounding discharges	41	37
Natural differences	2	2
Difference in physical habitat	15	13
Effluent-related effects?	32	28

The most common problems were with confounded discharges, incomplete information and differences in physical habitat. A closer examination of these studies would be

needed in order to identify how the study design could be altered to address these problems.

Twenty-eight percent of the studies indicate that there may be mill-related effects. Insufficient information is available, however, to describe the spatial extent or magnitude of the differences and whether the differences are indeed related to mill effluent. This would require a more detailed examination of the Cycle 1 reports.

The above analysis serves to illustrate that, if the Regional review of Cycle 1 reports were available prior to the BCEWG starting its own review, the group could have taken a more quantitative approach in this report, for instance, by categorising mills according to the number which showed a) no differences; b) differences but the results were confounded; c) differences and no confounding factors were evident. From this information, the type and extent of problems could have been identified and prioritized. It would have also provided an opportunity to indicate where additional information was needed from the Cycle 1 reports, for example, to provide support for recommendations to change the study design in Cycle 2.

In the absence of the regional summary reviews, the remaining discussion of Cycle 1 in this report is restricted to the information available from the sources cited at the beginning of this section.

In Cycle 1, mills had the option of using either an extensive or intensive study design. The extensive design, orientated towards mills where no prior information existed, incorporated a minimum of 12 "stations" (e.g. 6 reference, 6 exposed) sampled in duplicate. The intensive survey required a minimum of sampling 4 "areas" in triplicate for each of the exposure and reference areas, and was applied in situations where some spatial description of effects had been defined from historical studies.

The known number of mills (74) using an intensive approach suggests that there was already information available on the spatial extent of impacts. However, it should be pointed out that mills could have qualified for intensive surveys regardless of whether or not historical studies demonstrated mill-related effects. More importantly, since Annex 1 described the intensive and extensive surveys in terms of "minimum" sampling requirements, there may have been confusion as to what type of study design was used in Cycle 1. Some examples from the Atlantic region illustrate this paradox where the classification of 'extensive' or 'intensive' may have been based, for instance, on the number of replicates:

Mill	# Exposure/Reference	# Replicates	Study Designation	Alternate Designation
A	13/16	2	extensive	extensive
B	6/3	3	intensive	hybrid
C	12/10	3	intensive	extensive

Thus, inferences of spatial extent of "effects" cannot necessarily be inferred by the study design used in Cycle 1.

Most of the studies did not clearly map or delineate the extent of bottom sediment and faunal impact from the effluent source. Sampling protocols recommended in the technical guidance document were inadequate to do this. Since this was one of the original requirements of the monitoring program, the Cycle 2 requirements should clearly address this problem.

A specific observation from the marine studies was that there were too few "exposed" stations included in most studies to properly delimit the spatial extent of the pulp and paper mill impact. Furthermore, too few or inappropriate "reference" stations were used in some studies to allow proper comparison with exposed ones.

3.3 Confounding Factors

A number of factors from the Cycle 1 results have been identified that confound the interpretation both at the scale of absolute effects at individual mills and relative effects of mills across watersheds, regions or the country. Local confounding factors include natural ecosystem perturbations (e.g., in the marine environment: storms, which can decimate a benthic community), historic depositions, multiple discharges and lack of appropriate reference sites. On a larger scale, differences in sampling methods confounded between-mill comparisons.

3.3.1 Historical impacts

Methods employed to monitor benthic invertebrates in Cycle 1 were often insufficient to distinguish effects of past events from present discharges, so that it could not be determined if current effluent standards are adequate. For example, more than 38% (10 of 26) of the Ontario studies identified historical deposits of material as confounding the interpretation of current effects from mill discharges.

3.3.2 Multiple discharges

The EEM programme was designed to assess the impact of point-source discharges on the receiving environment. However, the approach was problematic in situations where other point (or non-point) sources of contamination were present. In Ontario, for example, 14 of 26 mills were influenced by other discharges. A further problem was that methods employed in Cycle 1 could not distinguish between organic inputs from pulp mills and other sources (e.g., municipal discharges). In order to assess these issues it will be necessary to consider the suitability of site selection and supplementary field

measurements so that better circumstantial evidence supports the proposition that the pulp mill effluent is responsible for the community changes determined in the receiving environment. Cause-effect relationships can only be established by experimental techniques (e.g., mesocosms) or analytical chemistry (pulp mill chemical identified in biota).

3.3.3 Reference sites

The selection of reference (or control) sites was another important issue in Cycle 1. In addition to being subject to exposure to other effluents a substantial number of reference sites were confounded because of habitat differences with the exposed sites (Ontario region, Quebec region, Atlantic region, Pacific Region).

3.3.4 Sampling Methods

Sampling methods used in Cycle 1 for pulp and paper mills in the Pacific/Yukon, the Prairies, Ontario, and the Atlantic regions were reviewed. The following methods were included: (1) sampling devices and their sizes; (2) mesh sizes used in the field and lab; (3) and sub-sampling methods. An attempt was also made to determine whether habitat sampled in coastal regions was freshwater or marine. A summary of the regions for which data were available follows.

In the Pacific region, marine and estuarine subtidal and intertidal soft-bottoms were sampled with a ponar grab (6 mills) or ponar/Smith-McIntyre grab combination (5 mills). Some studies used both samplers in the same survey due to sampler malfunctioning (Smith-McIntyre), logistical timing or physical constraints of sampling area (e.g., shallow depth). The Smith-McIntyre grab collected twice the area (0.1 m^2) of the ponar (0.052 m^2). Two mills sampled intertidal erosional habitats by artificial *Fucus* baskets over a 5-week period. One mill assessed the same habitat type using 1 m^2 quadrats. Quantitative dive transect surveys were undertaken at 2 mills to assess subtidal rocky shoreline habitats (one vertical, one horizontal). One included a video recording as well as visual profile observations. Erosional river habitats were sampled by a 35 cm dia. T-sampler (1 mill), a modified 35.6-cm diameter Hess sampler (5 mills), or benthic trays (2 mills). The colonization time for the artificial substrates ranged from 4-6 weeks. Depositional areas in rivers were sampled with an 15.2x15.2 cm Ekman grab (7 mills). Depositional areas in lake/reservoirs were sampled by a 0.052 m^2 ponar (3 mills). Screen sizes used to sieve subtidal and intertidal depositional samples were 180 µm, 250 µm, 500 µm, and 1 mm. The 180 µm effectively collected most of the meiofauna and was often used for historical consistency. Freshwater depositional samples were sieved with 180 µm and 250 µm screens. The finer fraction (< 1 mm) of most replicates was subsampled (20 of 23 mills) the >1 mm fraction was counted in full. The degree of subsampling ranged from 1-8 times (1/2 to 1/132). Limited information was provided on the method of evenness and accuracy of sample splitting. Two mills used a Folsom splitter. A few mills reported accuracies; one of these mills noted potential differences up to 24%.

Reports for the 10 mills in the Prairie region were reviewed. Seven mills were in Alberta, 2 in Manitoba, and 1 in Saskatchewan. All were located on rivers. Five different devices or combinations of devices were used; these were very dependent on the consultant who did the study. For example, one consultant used the Neill-Hess cylinder for their 4 mills; a second consultant used the Neill cylinder in erosional zones and an Ekman grab in depositional zones for their 3 mills. The sizes of devices were reasonably consistent. The Neill-Hess cylinders used by the first consultant were all 0.0892 m^2, whereas the second consultant's Neill cylinders were 0.093 and 0.1 m^2 (2 times). Other sizes: a Ponar and a mini-Ponar were both listed as 0.052 m^2, a Hess sampler was 0.09 m^2, and Ekman grabs were 0.065 and 0.052 m^2 (3 times). Mesh sizes were also consultant-specific and highly variable, ranging from 180-500 µm. For example, one consultant used 210 µm in their Neill-Hess cylinders in the field, but 180 µm for sieving in the lab. A second consultant used 210 and 200 µm in the field and lab respectively for erosional sites, and 500 and 200 µm in the field and lab respectively for depositional sites. The other 3 studies were highly variable; only one did not report (lab) mesh size. Sub-sampling was done for 9 of 10 mills. Of the 9, 2 did not describe their methods. One consultant used the Wrona *et al.* (1982) method for their 4 mills. Another consultant used a Folsom splitter for 2 of their mills and referred to a gridded-Petri-plate method (Cuffney *et al.*, 1993) for the third. Efficiency tests were not reported.

In Ontario, information on sampling methods was available for 22 mills. No data were available on sub-sampling. Nine different devices or combinations of devices were reported. However, the Ponar was by far the most commonly used device (11 mills). The size was usually reported as 0.052 m^2. A variety of other grabs was also used (Shipek, mini-Ponar, Ekman, Hester-Dendy artificial substrate sampler). Some mills used a mix-and-match approach to sampling devices. For example, one mill used 3 different grabs: a Ponar in the reference zone, a mini-Ponar in the near field, and an Ekman in the far field. Another mill used an airlift in the near reference and near field and a Ponar in the far reference and far field. Twenty out of 22 mills used 500 µm mesh for field operations; lab mesh size was not identified in the Ontario spreadsheets.

In Quebec, most mills used a Ponar grab in depositional areas and a Surber sampler in erosional habitats. A few mills used artificial substrates. The mesh size used was 500 µm. Subsampling procedures were used but not always described.

In the Atlantic region, information was taken from the 12 reports made available before the first meeting of the BCEWG and another 5 mills were added from Environment Canada spreadsheets, for a total of 17 mills. In cases of discrepancies, the reports were used. The spreadsheets had no information on sub-sampling. The mills had the following habitat distributions: 3 freshwater, 4 freshwater and estuarine, 3 estuarine, and 7 marine. Five different sampling devices or combinations of devices were reported. By far the most common were the van Veen (8 times) and Ponar (6 times) grabs. The size of the van Veen was variously reported as 0.07 m^2, 0.06 m^2, and 0.1 m^2 (and not reported twice). The 0.1 m^2 size was probably a rounded figure for 0.06 m^2, which seems to be the

exact size. The Ponar was always reported as 0.05 m^2 (but size was not given once). The Ponar and van Veen were used in both freshwater and marine habitats, according to the spreadsheets. The Ekman was used only once (in fresh water) in combination with a van Veen (marine). A 500 µm mesh size was used uniformly, but field or lab use was not distinguished. Of the 12 mills for which data exist, no subsampling was done for 3, it was done but not specified for 4, and it was not reported for 5.

Regional reviews and workshops suggested that standardised methods should be considered an advantage as it would allow Environment Canada and industry to make regional and national assessments of the environmental performance of mills.. At present, the department has in hand many individual mill studies but larger-scale examination can only be done qualitatively. A possible disadvantage to standardising protocols (EVS consultants in their Pacific summary) is the loss of historic comparability. However this can be addressed by conducting parallel studies with old and new methods for a survey. We suggest the long-term gains at local (cost effectiveness of the new methods) and regional scales outweigh the benefit of retaining old methods.

3.4 Data analysis

There was a major emphasis in Cycle 1 on the use of multivariate analyses for assessment of effects. Multivariate methods may be appropriate when a multiple-control-impact or radial-gradient approach is used. Generally there was inadequate or inappropriate statistical analysis of data from the benthic macroinvertebrate sampling. The suggested recommendations for data analysis (provided in Section 4.3) are based on discussion among the BCEWG members, a previous review of the EEM guidelines and consultant's reports from Cycle 1 studies.

One region suggested that a standardised data management system be considered, to allow meta-assessments to be conducted. As part of its review of Cycle 1 electronic reporting, the data management expert working group has provided recommendations on a standardized electronic data structure and reporting system that will be implemented during Cycle 2.

Other aspects of data interpretation were raised in regional reviews. The first was related to the necessity of using complex data analyses, specifically the use of correspondence analysis ordination followed by ANOVA of ordination scores (Alberta Forest Products Association [APFA] workshop, Ontario reviews, Atlantic region). The general view of many reviewers was that simpler univariate analyses and indices would be equally useful for this type of study design with a small number of sites (AFPA, Ontario reviewers, Quebec region, Atlantic region). The second interpretative issue was related to what constitutes an ecologically significant (as opposed to a statistically significant) effect. This was not clear in most of the reports. What constitutes an ecologically significant effect must be incorporated into the study design before the studies are conducted; the

existing Cycle 1 data will be valuable in determining this on a site specific basis (see Table 6.3 and Section 4.3). Finally, several regions suggested there was a strong need to link physical and chemical data with the invertebrate data (Council of Forest Industries [COFI] workshop, AFPA workshop) so that cause-effect relationships could be elucidated (IOS workshop and COFI workshop, AFPA workshop, Quebec region, Atlantic region).

3.5 Summary

The Cycle 1 results for benthic macrofaunal surveys went a long way towards meeting the overall objectives of the EEM programme. Most of the consultant's reports examined indicated a high degree of expertise and precision in the handling of field samples. Identifying and characterising the habitats at risk in marine environments has been greatly facilitated by the results from Cycle 1 surveys. In addition, potential confounding factors and sampling problems for each mill area have been identified.

The Cycle 1 design was preliminary and theoretical, and was intended to provide sufficient data to characterise conditions at all the mills and help the design of Cycle 2. Cycle 1 was a success in these goals due to the general high quality of the reports. The major observation of the BCEWG from Cycle 1 is that invertebrate community structure surveys can delineate differences between reference and exposed areas. Of the 112 individual mill studies examined for this report 69% showed differences between reference and exposed sites. However, in a large number of cases, it was not possible to relate the observed changes to mill effluent because interpretation of the data was confounded by factors such as historical effects, multiple discharges, and the location of the sampling sites. Moreover, because of the variety of methods used, only qualitative comparisons among mills could be made. Finally, it was established that further guidance on appropriate data-analysis techniques is required.

Table 3.1. Summary of issues identified regionally

Concern	Pacific	Prairie & Northern	Ontario	Quebec	Atlantic	Total
Site selection issues						
Inflexibility of design	x	x	x			3
Locations	x		x	x	x	4
Habitat type	x		x			2
Reference sites	x		x		x	3
Confounding factors	x		x	x	x	4
Methodological issues						
Protocols	x		x			2
Sample methods	x		x			2
Core variables	x					1
Taxonomic resolution	x		x			2
Replication	x					1
Interpretation issues						
Data management	x					1
Data Analysis	x		x	x		3
Decision criteria - effect size	x		x	x		3
Other						
National consistency	x					1
Regional approach	x					1
Inclusion of other data	x					1

The view of the BCEWG is that Cycle 2 should focus on resolving issues related to confounding factors, providing advice on standardised methods where appropriate and making the data analysis more relevant and appropriate.

These conclusions regarding the results of Cycle 1 are similar to those made by various regional reviews. The comments of regional co-ordinators fall into three major areas: site selection, methods and data interpretation (Table 3.1).

4.0 Recommendations for Cycle 2

The results of Cycle 1 did not provide a basis for comparing national or regional impacts or effects in receiving environments because no guidelines were provided by which sampling programs could make such a comparison. Based on the BCEWG's review of Cycle 1, the ability to relate changes to mill discharges, the suitability of methods for data analysis and problems associated with standardisation of sampling methods were the major issues. These can be grouped into three broad categories (site selection, sampling methods and data interpretation), which form the basis of the BCEWG's recommendations for developing Cycle 2.

Table 4.1: Benthic sampling programme designs

Design Type	Reference / "control" Sites	"Impact" Sites	Statistics	Considerations
Control/Impact (CI) or	Single	Near field/far-field/far-far field as defined by effluent concentration in receiving environment	ANOVA	Is the single reference site confounded by other environmental factors?
Linear gradient	No or low exposure stations	" "	Regression/ multivariate	Can suitable no- / low-exposure sites be found?
Multiple - CI (MCI) or	Multiple sites in the same or environmentally similar adjacent watersheds/coastal areas	Near field/far-field/far-far field as defined by effluent concentration in receiving environment	Multivariate	Multiple "reference" locations
Radial gradient	Replicate no- or low-exposure stations	" "		

4.1 Site Selection

The recommended sampling program designs for the freshwater and marine benthic community assessments for Cycle 2 are summarised in Table 4.1. These designs are based on an evolution of the approaches used in Cycle 1. Each design is aimed at

estimating the magnitude and spatial extent of a potential effect related to pulp-mill effluent, taking into consideration factors and possible constraints related to the availability and spatial distribution of suitable control ("reference") sites and the spatial extent and number of potential impact sites (e.g., near field/far-field locations).

The simplest study design used in Cycle 1 environmental effects monitoring is the control-impact (or reference-exposure) design (subsequently referred to as the CI design) (Green, 1979) (Figure 4.1). In rivers this consists of one upstream reference site and a series of downstream exposure sites representing near-field, far-field and far-far-field sites. In lakes, one reference area is typically selected in a nearby bay of the same lake and exposure sites are selected at increasing distances away from the mill effluent. In certain marine and freshwater environments it is not possible to define a control or reference site immediately upstream from the discharge of the mill effluent. In such cases a modified CI, gradient approach is used; sampling sites are selected along a gradient of effluent exposure. In this situation, reference sites (or control sites) are designated as those with the least mill influence (e.g., a set of far-field locations with low or no exposure would constitute the reference or control condition for comparison).

Two of the major problems associated with the use of a single reference site are that it can be easily confounded by other factors and it is pseudoreplicated (Hurlbert, 1984). For example, a single reference site may differ from an exposure site with respect to many environmental variables in addition to the point-source effluent. If such circumstances occur, a Multiple-CI or radial gradient design should be used (Table 4.1, Figure 4.2). Reference or control sites should be similar in as many traits as possible to exposure sites.

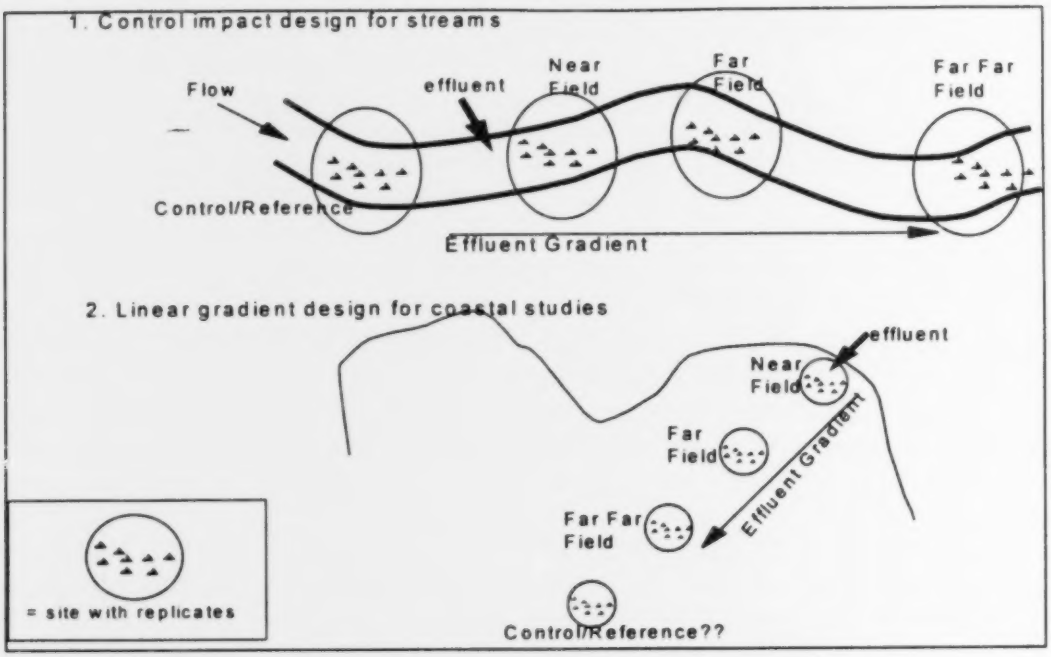


Figure 4.1 Control/impact and linear-gradient designs

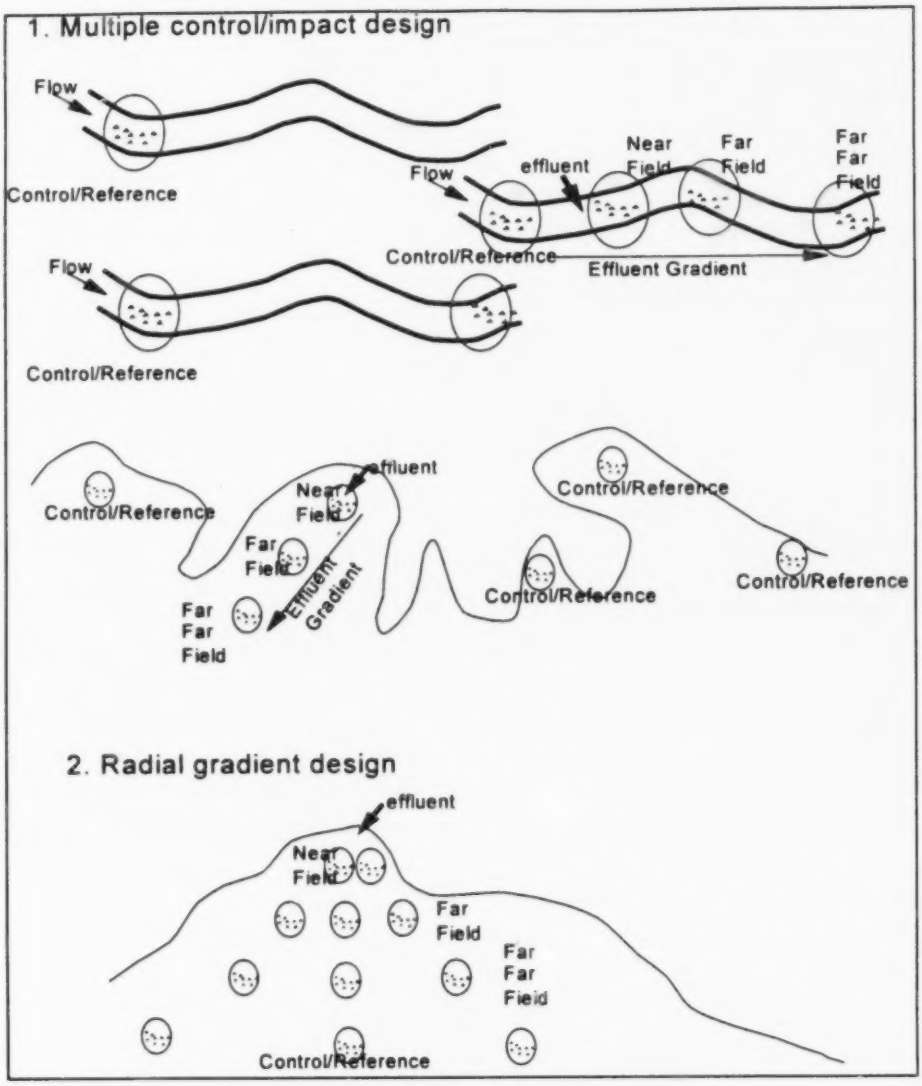


Figure 4.2: Multiple refernce site designs

- *It is recommended that a protocol for the a priori selection of reference sites be used.* If single reference sites in extremely disturbed areas are impossible to locate (i.e., the simple CI or linear gradient design is not appropriate), the multiple CI or radial gradient approach should be **used.** Sampling schemes should be devised so that reference sites with minimal degradation are located in comparable habitats within the same ecoregion. For coastal or lake-situated mills, a gradient or radial-gradient sampling approach away from the effluent source may be used depending on site-specific circumstances. This method assumes that reference information is collected from the stations along the gradient with the least exposure to the effluent source. A visual survey of the affected area around the mill will help to

delineate affected areas and to identify confounding factors from other organic inputs influencing the seabed in the area of the mill.

- *It is recommended that additional reference sites should be included in future cycles so that among-area variation can be quantified and exposure-area effects can be evaluated relative to reference site variation.*

  Alternatively, a data set encompassing reference areas from multiple EEM surveys conducted in the same season within the same ecoregion could be used. Several benefits could be realised through co-operation among mills in monitoring. Substantial cost savings could result for mills located within the same drainage basin through the co-ordination of sampling sites, sharing of field equipment, standardisation of field gear and data processing. Such co-ordination may reduce the effects of confounding factors and so more accurately quantify mill effects.

- *It is recommended that for those mills in areas where reference data bases (Fraser River basin, Great Lakes, Puget Sound) have been established, that methods of data collection allow use of the reference sites in the databases for site assessment.*

  This, of course, would require that the approach be validated using the existing reference data base and the results from Cycle 1.

- *It is recommended that a standard nomenclature should be used.*

  It was often difficult to discern what constituted a replicate, site, station, area or zone. Most companies had difficulty in selecting appropriate independent sampling locations and confusion existed in distinguishing between replicate samples at a site and independent sites. Accordingly, the study design and subsequent reports should include clear operational definitions of the following terms: replicate, site, habitat. It is recommended that standard definitions be included in future revisions to the Technical/Further Guidance Documents. For example, a site should represent a reach of river (usually a distance 6 times the stream width) or a location in a lake or marine environment with re-visitable dimensions (10 m x 10 m = 0.01 ha) from which a number of replicates are taken. These replicates represent the sample.

- *It is recommended that site selection should incorporate the requirement to determine spatial extent of effects from the effluent source.*

  Sufficient site replication must occur in the design to allow for near-field, far-field and far- far-field comparisons with reference or control conditions.

- *It is recommended that the criteria for determining the number of replicates be standardised.*

  The number of replicates should be determined at each site such that statistical tests can detect the required amount of difference between sites of different exposure to effluent.

4.2 Sampling Methods

- *It is recommended that only the dominant habitat type be sampled at a site.*

  In Cycle 1 it was a requirement that habitats be sampled proportionally according to their occurrence. For purposes of cost effectiveness and ease of data interpretation, the BCEWG recommends that a decision be made to sample either depositional or erosional zones in the freshwater receiving environment. The sampling of both zones is probably not necessary for adequate biomonitoring. In addition, decisions about sampling intertidal versus subtidal substrates in the estuarine/marine mills will depend on which is the appropriate receiving environment.

- *It is recommended that standard sampling devices be used as much as possible.*

  Depositional zones in freshwater systems should be sampled by either Ponar or Ekman grab. Only the Atlantic region used the van Veen grab with regularity. If no particular reason exists for this use, then the Ponar or Ekman should be used there, too.

  Erosional habitats in freshwater environments can be sampled by the Neill-Hess cylinder which will allow unit area (0.1 m^2) estimates to be made. However, a more universal use of kick-net samplers should be considered because they are simpler to construct, more versatile, and less expensive than the Neill-Hess samplers. A 3-min, standard time collection should be adequate.

  Depositional habitats in marine environments can be sampled with the Smith-McIntyre grab, a modified Van Veen grab, or mini (petite)-Ponar grab which are suitable and available in Canada. Regardless of the sampler used, the preferred unit area sampled is 0.1 m^2. A good review of marine sampling methods is available in Holme and McIntyre (1984).

  Erosional habitats in marine/estuarine environments (intertidal and subtidal) should be sampled using a standard 0.1 m^2 quadrat.

  The use of artificial substrates and mixtures of sampling devices should be avoided unless there are no other alternatives available. Artificial substrates require at least two visits to take one sample and, unless they mimic the natural substrate, attract an uncharacteristic assemblage of invertebrates.

- *It is recommended that consistent mesh sizes be used for processing samples in order to allow better evaluation and comparison of mill performance regionally and nationally.*

  All freshwater mills should adopt 500 µm mesh for field and lab use. Currently, 500 µm mesh is overwhelmingly used in Quebec, Ontario and Atlantic regions, but a great variability exists among mills and often between field and lab use

elsewhere in Canada. The larger-sized mesh will reduce processing requirements for the mills, and offer some degree of standardisation among regions.

All marine studies should use a stacked set of 0.5 and 1 mm screens in the field, with 0.5 mm samples archived and only processed if necessary to obtain required power for statistical tests. Data should be analysed and recorded separately for 1 mm samples nationwide to standardize results. This will provide the maximum information and reduce sorting and taxonomic costs. As an adjunct, adult abundances should be recorded and analysed separately to avoid confounding from recent recruitment.

- *It is recommended that sorting, sub-sampling and taxonomic efficiency must be standardised as much as possible and must be clearly documented for each mill.*

  For large samples, processing time adds a substantial cost. Sub-sampling has been shown to be a reliable and effective approach to addressing this problem when used appropriately and with the required QA/QC. In Cycle 1 sub-sampling was often not reported or described. The descriptions that were available indicate a variety of methods were used, but the efficiency was rarely reported. The degree and frequency of sub-sampling should be reported in the text and in the data tables. Specifically, sub-sampling information should be included in the row entry and in each data column of tables that report numbers of organisms in samples.

4.3 Data Interpretation

4.3.1 Supporting Field Measurements

- *It is recommended that a standard set of physical and chemical habitat attributes be measured.*

  The measurement of environmental attributes can discriminate changes between natural and mill-related effects and relate observed effects to mill discharges. We would suggest that the variables listed in Table 4.2 be considered for inclusion in the sampling programme to assist in the interpretation of data from invertebrate community surveys.

Other good sources of information on selecting habitat descriptors for explaining patterns in lotic invertebrate communities are Wright *et al.* (1984), Corkum and Currie (1987), Corkum (1991) and Ormerod and Edwards (1987). For lentic communities, papers by Johnson *et al.* (1991), Reynoldson *et al.* (1995) and Bailey *et al.* (1995) provide lists of the most useful habitat descriptors. Several authors (Clarke, 1993; Clarke and Ainsworth, 1993; Hargrave *et al.*, 1993, In press; Warwick, 1993; and Warwick and Clarke, 1993) have described responses in invertebrate communities and demonstrated the ability to distinguish natural variation from anthropogenic stressors. These authors should be consulted when developing a list of the most appropriate descriptors to measure. Some of these variables will be required and others will be optional as described in the Technical Guidance Document.

Table 4.2 Variables for freshwater (F) and marine (M) habitats

VARIABLE	HABITAT (F, M)	JUSTIFICATON
Soluble reactive or total dissolved phosphorus and total phosphorus	F	Phosphorus is often the limiting nutrient in fresh water. Mills often discharge large quantities of P, which leads to nutrient enrichment.
Nitrate-nitrite, ammonia, and total nitrogen	F	Nitrogen is often a secondary limiting nutrient in fresh water. High ammonia levels can lead to toxic conditions for fish.
Dissolved organic and particulate carbon.	F	Carbon is a nutrient source for microbes. Mills discharge large quantities of carbon sources.
Sodium	F	Na is a good tracer for bleached kraft mill effluents.
Algal biomass as chlorophyll or ash-free dry mass	F	Algae in the water column (lakes) or on benthic substrates provides a food source for higher trophic levels.
Dissolved oxygen near bottom.	F, M	Dissolved oxygen can be decreased by mill effluents due to high BOD.
Water temperature	F	Mill effluents can cause substantial increases in water temperature in receiving waters.
Salinity	M	Changes in estuarine conditions profoundly affect benthic communities.
Colour or turbidity	F, M	Many older mills discharge effluents that are deeply coloured. This may reduce the light available for primary production.
C/N ratio	M	C/N ratio is affected by terrestrial (wood) versus plant sources of marine deposition
Total organic carbon or loss on ignition	F, M	Effluent effects are often related to inputs from organic material, which allows relationships between response and enrichment to be established
Bed structure - particle size, matrix: framework ratio, degree of embeddedness, texture, colour and thickness of layers in cores	F, M	Differences in physical structure of the habitat can influence invertebrate community structure. One approach to ensuring reference and exposed sites are equivalent is by measurement of the bed structure. Core data help separate historical from present-day effects.
Current velocity	F, M	Provides information on equivalence of sampling sites and eliminates one source of confounding interpretation.
Depth	F, M	Water depth has a major effect on invertebrate communities.
Bankfull channel width	F	Again a good surrogate for habitat structure in stream environments.
Ph	F	Can eliminate confounding issues and provide basic information on water quality.
Alkalinity	F	Provides a good surrogate for water quality.
Conductivity	F	Provides a good surrogate for water quality.
Eh (redox)	M	Sediment redox provides an indication of the anoxic/oxic boundary in sediments
Total sulfides	M	Sulfides in marine sediments indicate the extent and nature of microbial response to organic enrichment.
Latitude, longitude	F, M	Provides basic information to ensure site integrity in temporal data series.
Chemical tracers of mill effluent	F, M	Provide evidence of spatial extent of plume.

4.3.2 Data Measurement Variables and Analysis

The operative word here is "flexibility" since the message was very clear from Cycle 1 that the data-analysis requirements were overly complex and restrictive. The data-analysis methods recommended here are suggested as a minimum for reporting purposes and to provide some consistency among investigators. It is recognized that there are a variety of methods for analyzing data describing benthic invertebrate communities. Therefore, the emphasis here is as much on graphical and tabular reporting of temporal and spatial trends in the communities, along with simple, documented, and accessible database creation and storage. If raw data are stored in a safe, accessible manner, then any re-analysis or meta-analysis considered appropriate can be done with minimal cost and maximum efficiency.

The comparison of (perhaps several) reference and (various degrees of) exposed sites in the context of a single assessment study is straightforward. Although there are statistical limitations to this approach (pseudoreplication), the combination of a clear graphical description of the spatial pattern of community descriptors with a one-way Analysis of Variance (ANOVA) or Analysis of Covariance (ANCOVA) will illustrate any site-specific effects of the effluent on the benthic invertebrate community.

Most commonly, the presence or absence, proportion, or density of each benthic invertebrate taxon in the community is the basic information obtained in any benthic survey. Such data form "the fundamental dataset" and are recommended variables to include in the EEM benthic invertebrate community survey. In marine systems, some measure of relative biomass can be useful. The following describes how the fundamental dataset can be used:

- *It is recommended that community descriptors such as taxonomic richness, diversity and equitability are all functions of the variables in the fundamental dataset. Calculations, statistical analyses and particularly graphical presentations of these may aid in the detection and description of the effects of exposure.*

  Species richness and diversity indices are only useful if most of the fauna have been identified to species. If not identified to species level then only family, genera or other taxonomic group richness can be reported. Alternative classifications (e.g., sensitivity indices, functional group proportions or densities) can be also be used to re-express the data if sufficient independent information about ecological tolerances or preferences of community members is available. Calculations, statistical analyses and graphical presentations of these may aid in the detection and description of the effects of exposure.

  Similarity indices such as the multivariate distance of the community from the "average" (median or mean) community can be calculated for each observation using any of a number of distance measures (e.g., Bray-Curtis distance). Such

distances are the raw material for the ordinations and cluster analyses that some researchers feel are particularly valuable in identifying patterns in community variation among exposure areas. Calculations, statistical analyses and particularly graphical presentations of community distance measures may aid in the detection and description of the effects of exposure.

- *It is recommended that communities be described at both the individual-observation and site spatial scales.*

  For example, taxonomic richness can be expressed either as an average per individual observation at a site or as the total number of taxa observed at a site. The same is true of other community descriptors listed above. Descriptions at both spatial scales are potentially informative. Although descriptions at the site scale are not amenable to statistical analysis within the context of one study (because of a lack of replication), meta-analysis of data from several studies could statistically test for patterns in such descriptors.

- *It is recommended that for marine samples a simple, comparative measure of macrofaunal biomass would be useful to compare and contrast sediment quality conditions, since the larger taxa now appear to be the best indicators of impact.*

4.3.3 Data handling and screening

- *It is recommended that the distribution of variables from observations within exposure areas (e.g., reference, exposed) should be examined for normality and equality of variance if parametric hypothesis tests are to be carried out.*

  Invertebrate surveys result in either absolute or proportional density or biomass estimates. Both the fundamental dataset, as described above, and descriptors that are either functions or re-expressions of the original variables need to be screened for transcription and calculation errors.

- *It is recommended that once the data have been screened for coding or counting errors, all data should be stored in simple, accessible databases.*

  Visual screening techniques such as box and whisker plots, and stem-and-leaf diagrams can be used to identify the general nature of the distribution of variables as well as extreme values. This should be done for both variables in the fundamental dataset and calculated community descriptors.. The complete fundamental dataset, including rare and highly variable taxa and ambiguous identifications should be stored in this manner, even if data filtering has been applied prior to calculation of community descriptors. This way, alternative approaches to data filtering, calculation of community descriptors, and analysis can be employed in re-analysis or meta-analysis.

- *It is recommended that data should be entered on a standard data base (e.g., Microsoft Access), to facilitate comparisons of data sets among mills.*

  This is being addressed separately by Ottawa under an information management framework. Data need to be reported in a straight-forward, readily useable computerized format for larger-scale comparisons by regulators.

- *It is recommended that values for means, standard errors and simple statistical tests be included in future reports.*

  Although raw data sheets were provided in all reports, values for means, standard errors and basic statistical tests for differences in basic community descriptors among sites were lacking. Future cycles should include these data and analyses.

4.3.4 Data manipulation

- *It is recommended that organisms that cannot be identified to the desired level of taxonomic precision should be reported as a separate category in the fundamental dataset at the finest level of taxonomic resolution possible.*

  The investigator can make a reported decision as to whether or not to include these taxa in the calculation of community descriptors. The data are clearly identified so that a different decision can be made on re-analysis or meta-analysis of the data.

- *It is recommended that data transformations be used if they are appropriate for statistical reasons such as stabilising of variance for parametric testing.*

  This re-expression describes proportional rather than absolute variation among communities. The scale of application of such a transformation should be clearly stated (e.g., sub-sample numbers transformed prior to averaging, or averages from sub-samples transformed). Still, the un-transformed data should be stored in accessible fashion to facilitate alternative transformations or analyses.

4.3.5 Effect size

In order to answer the question as to whether current effluent regulations are adequate to protect the environment, and whether benthic community surveys have adequate power to detect effluent effects, some decision criteria need to be set on which to base management decisions. In Cycle 1, differences in benthic communities exposed to pulp mill effluent included a reduction in the number of species, changes in abundance, presence of organic enrichment indicator species and the absence of sensitive species. Effect size, the magnitude of relationship between the benthic community and the effluent that is deemed ecologically significant, should be defined for each community (Table 6.3). If the test of a particular community descriptor in a study has less than 80% power (chance of detecting

an effect if one is truly present), then the study is inadequate to conclude that a change in a given community descriptor is unrelated to effluent exposure.

As outlined in Table 6.3, a promising method of setting an ecologically relevant effect size is to define it in terms of the normal range of variability of a particular community descriptor in relatively non-impacted reference communities. Further detail is available in Lowell (1997) and the approach is briefly summarized here. To begin with, a widely accepted range that could be used for EEM is ±2 standard deviations (±2 SD) calculated from the reference communities; that is, the 95% confidence interval for the reference communities. For some mills, initial estimates of reference community variability are too high to be of use for setting effect size. Therefore, it is recommended that standardized habitat-specific effect sizes be calculated by taking the medians of ±2 SD across several mills within given regions. For example, for the common habitat types sampled in the Pacific + Yukon region, initial estimates of median effect size for number of species range from ±15% to ±60%.

These initial estimates will need to be revised, however, as more data are collected during subsequent EEM cycles, because initial variability estimates were based on the very low numbers of reference stations sampled per mill (usually 2 or 3) during Cycle 1. Therefore, the number of reference (and impact) stations should be increased in subsequent cycles. The needed degree of replication cannot be provided by subsamples within stations because subsampling variation is at the wrong spatial scale for tests of area-wide effluent impacts and effectively amounts to "pseudoreplication". Using the ±2 SD criterion, recommended sample sizes (number of sampling stations per reference and impact areas) can be calculated by rearranging the appropriate power analysis equation. For a simple reference versus impact area(s) comparison, 5 sampling stations per area would be required to achieve the recommended minimum power of 80%. Many workers suggest that power should be set so that both Type I and Type II errors are equal to the 0.05 probability level (i.e., power=95%). To achieve 95% power, 8 sampling stations per area would be required.

An additional caveat is that effect sizes that are initially set based on reference area variability may need to be adjusted up or down based on any *external ecological information* that may become available about the nature of response to the effluent expected for a particular community descriptor (Table 6.3). Additional adjustments may also be required based on *societal values* regarding what is considered to be an important change in the ecosystem. Finally, although initial discussion of effect sizes has emphasized potential changes in average community conditions, effluents may also have negative impacts on *community variability* itself.

Any possible criteria for sediment standards that can be applied to the mills environment based on existing, well-researched objectives should be tested in Cycle 2 to examine their performance.

4.4 Technical Workshop

It is recommended that a technical workshop be established for EEM in the three habitats (riverine, lacustrine, marine/estuarine) in which mills are located. The workshops would ensure consistency and standardization of all aspects of EEM. The workshops would also provide an opportunity for representatives from industry and consulting companies to work co-operatively in the standardization of techniques. The workshops could be co-ordinated by representatives of industry, Environment Canada, academics and consultants. The focus of the technical workshops would include:

- study design (site selection, replication);
- sample collection (field component: biotic and environmental attributes; sampling protocols);
- sample processing (laboratory component: QA/QC, subsampling, sorting, identification)
- data analysis and interpretation (use of sample data sets).

5.0 Available Alternative Methodologies and Approaches

The work group was requested to provide information on the existence of alternate approaches to identifying impacts on invertebrate community structure. While at the present time no alternate approved methods exist, a number of approaches should be considered for testing and approval, where situations warrant. These methods are described more fully in Section 7 and could be an important component of future EEM cycles.

6.0 Decision Tree

The work group was also requested to develop a decision tree to assist in the design of Cycle 2 studies. A decision tree has been developed through several iterations for application in both freshwater and marine study designs (Fig. 6.1). This together with the accompanying tables should provide the basis for determining the frequency and extent of studies through several EEM cycles.

An earlier (but similar) version of the decision tree was tested for 53 mills in the Ontario and Quebec Regions. The outcome of these tests was that for the vast majority (> 85%) of mills, the statistical and sampling design should be reviewed or revised prior to undertaking the next benthic survey. In most cases, confounding factors were thought to

be resolvable with the application of a revised design. However, several mills had situations with unresolvable confounding factors; for these mills it was concluded that alternative approaches to benthic surveys would be needed

Figure 6.1. Benthic Survey Decision Tree

1
Did the benthic community differ between exposed and reference locations or was there a response to an exposure gradient (Table 6.1)?
No
Yes
2
Are there confounding factors potentially influencing the results (Table 6.2)?
3
Was the sampling design adequate to determine that effects were mill related (Tables 6.2 and 6.3)?
No
Yes
No
Yes
4
Was the power of the statistical test adequate to determine that there was no mill effect (Table 6.3)?
5
Can confounding factors be resolved by redesigning the benthic survey?
6
Was the sampling design adequate to determine the spatial extent of the effect (Table 6.5)?
Yes
No
Yes
No
Yes
7
Have either the mill effluent quality or quantity degraded since the last cycle (Table 6.4)?
8
Are approved alternative approaches available (Table 6.6)?
9
Modify sampling and statistical design to better define how the benthic community varies in relation to exposure to the effluent (Table 6.3).
10
Review next step with RAO
No
Yes
No
No
Yes
11
Is there an ongoing research program that satisfies EEM requirements or is there recent data for a conclusion and mill process and treatment have not changed?
12
No benthic survey in next cycle
13
Apply approved alternative approaches
No
Yes
14
Review sampling and statistical design and proceed to next cycle
15
No survey this cycle

Table 6.1: Did the benthic community differ between exposed and reference locations or was there a response to an exposure gradient?

Answering this question is equivalent to comparing various descriptors of the benthic community (e.g., total abundance or biomass, taxonomic richness) among the two or more levels of exposure to the effluent. A relationship between the benthic community and the effluent may be a simple difference in a community descriptor between one or more reference sites and an exposed site, or it may be a gradient of variability in a community descriptor among sites, which vary in their level of effluent exposure.
A weight-of-evidence decision must be made based on both statistical and non-statistical comparisons of sites with varying exposure levels.
Statistical comparisons are based on community descriptors calculated from each observation made at a site with a given level of exposure. An example is the average number of genera per replicate. Variation among sites with different exposure levels can be assessed with 1-way ANOVA or its non-parametric analogs. An ecologically significant effect size must be defined for each community descriptor subjected to this statistical test (Table 6.3).
Non-statistical comparisons are based on community descriptors calculated from the pooled observations made at a site with a given level of exposure. An example is the total number of genera found at a site. Variation among sites with different exposure levels can be assessed by inspection. If more than three levels of exposure have been sampled, the relationship between a community descriptor and exposure level can be statistically evaluated, albeit usually with weak power. Although statistical comparisons are usually not appropriate for these comparisons, an ecologically significant effect size must be defined for each community descriptor subjected to this comparison (Table 6.3).
A typical assessment, where several to many community descriptors are used, will result in a variety of judgments about the relationship between the benthic community and exposure to the effluent. For example, average number of genera per replicate may not be related to effluent exposure, but average total abundance per replicate may be. In the case of a mixed result, the previously established response of community descriptors to pulp mill effluent will be taken into consideration when constructing a weight-of-evidence decision. If descriptors known to be sensitive to the effluent are not correlated with exposure at a given mill, while other variables not known to be affected by pulp mill effluent are correlated, the argument for passing the site is strengthened.

Table 6.2: Was the design adequate to suggest that the effects were circumstantially related to pulp mill effluent? Are there confounding factors or sampling methods potentially influencing the results?

The relationship between the benthic community and the effluent may be augmented, reduced, or qualitatively changed by confounding factors or application of incorrect sampling methods. There are four categories of such factors:

a) confounded environmental descriptors.
b) multiple discharges or historical effects
c) time of sampling
d) sampling methods
e) natural causes of benthic disturbance (e.g., storm effects on sift sediment communities).

a) Confounded environmental descriptors.
If environmental descriptors (e.g., depth, substrata; see Table 4.2 in report for list of descriptors appropriate for different contexts) are compared among sites with varying levels of exposure to the effluent in the same manner as community descriptors (Table 6.1), and they differ in a manner consistent with the observed community variation, a potential for confounding effects exists. If feasible, a change in sampling design by modifying site locations at macro (e.g., geographic location), meso (e.g., riffle within a river site), and micro (e.g., substrata within a riffle) spatial scales may reduce the problem. Augmenting the design to better characterise reference conditions by sampling more reference sites (e.g., moving from a gradient to a radial gradient design, or a control-impact to a multiple-control-impact design; Table 4.1 in report) might reduce the problem.

b) *Multiple discharges or historical effects*
If sites with varying levels of exposure to the pulp mill effluent also have varying levels of exposure to other effluents or stresses, or habitat modifications from historical effects of the mill, a potential for confounding effects exists. If feasible, a change in sampling design by modifying site locations at macro (e.g., geographic location), meso (e.g., riffle within a river site), and micro (e.g., substrata within a riffle) spatial scales may reduce the problem. Alternatively, discharge fingerprinting techniques and mesocosm studies may better define specific effects of a given pulp mill. Core sampling may also be useful in depositional environments to resolve historical trends.

c) *Time of sampling*
The time of year or year of sampling, although more or less the same in all sites sampled, may confound the relationships observed by overlaying the relationship between the benthic community and the effluent with seasonal and multi-year temporal variation in benthic populations. This can be assessed by knowledge of the phenology of the members of the benthic community and examination of data collected in previous years from the reference sites.

d) *Sampling methods*
If sampling devices and processing of samples have been other than the standard recommended procedures any effect of the effluent on the benthic community could be augmented, reduced, or qualitatively changed.

Table 6.3: Was the power of the statistical tests adequate to detect a relationship between the benthic community and the effluent exposure if one was present?

If there is no relationship between a community descriptor and the effluent detected in a given study, the ***power of the statistical tests*** should be assessed. ***The power of a statistical test is the probability of detecting a given magnitude of relationship between the benthic community and the effluent***, known as the ***effect size***, when sites of varying levels of exposure to mill effluent are compared.

Effect size is the magnitude of the relationship between the benthic community and the effluent that is considered ecologically important to detect. A Type II error is the probability of not detecting a relationship (e.g., reference vs. near-field difference) from sampling in a given cycle if one of such a magnitude does, in fact, exist. A Type II error probability of 0.20 is typically used, although others have been considered.

Effect size of a particular community descriptor can be defined in three ways:

a) some function of ***variability*** in the community descriptor ***between observations at various exposure levels***
b) ***external ecological information*** about the acceptable magnitude and nature of response of a community descriptor to the effluent, *e.g.* critical species for commercial or scientific reasons.
c) ***societal input*** into what constitutes an important magnitude and direction of ecosystem change

The first approach to defining ***effect size*** quantifies the difference between an exposed community or communities relative to variability among those communities used for reference or comparison. Commonly, an observation outside of the range of 95% of the reference communities is considered odd enough to "fail". This method exploits variability among reference observations directly into the definition of power, but may identify irrelevant variables as powerful community descriptors. The density of a bivalve mollusc species may be relatively constant in Ontario reference streams, and it would seem to have potential as an extremely powerful test of the relationship between the benthic community and the effluent. However, if the mollusc is tolerant of pulp mill effluent, it would be a very poor community descriptor for bioassessment of pulp mills. Because of the nature of bioassessment studies, particularly their constraints with respect to cause-and-effect and confounding influences, community descriptors with low variability among reference communities but little known relevance to pulp mill effluents should be avoided.

The ***magnitude and nature of change we would like to detect in a bioassessment*** (***effect size***) should be some fraction of the known change in the community descriptor in response to pulp mill effluent. This, together with societal input, will define the effect size for various community descriptors. If a set of such community descriptors with known response to pulp mill effluent are determined, those with less variability among reference communities will be the most sensitive monitoring tools.

Once ***effect size*** is defined, the adequacy of the sampling design can be readily assessed. One simply asks "With the study design and sample size used, what was the power of detecting the ***effect size*** magnitude of relationship between the benthic community and the effluent?" If this was less than 80%, than the study was unable to conclusively show there was no relationship between community and effluent, even if the probability level associated with the statistical test was well over 0.05.

Table 6.4: Has either the mill effluent quality or quantity degraded since the last cycle?

The absence of a relationship between community structure and effluent is based on the relationship between the loading to and the conditions of the receiving system. If either of these changes **significantly** then the assumption of no impact is no longer valid, and a sampling programme should be initiated in the following cycle.

The following changes in effluent or the receiving system could trigger such a sampling programme depending on site-specific conditions:

A. **Effluent quality and quantity changes**:

1. Receiving system response is determined by the loading of contaminants (e.g., nutrients, toxics) to the receiving system. Contaminant loading is determined by both concentration and volume of effluent. Therefore, if either of these change and the RAO considers the changes to be potentially detrimental to the receiving system then a sampling programme is required (see below).

B. **Receiving system changes:**

The environmental effects on receiving systems are determined by the relationship between receiving system sensitivity and discharge quality and quantity. Therefore, if there is a change in the receiving system conditions (e.g., dilution capacity) and the RAO considers that the change makes the receiving environment potentially more sensitive to pulp mill discharges, then a sampling programme should be initiated. (See below.)

Guidelines are needed to assist RAOs in deciding whether changes in effluent quality and quantity and/or receiving system sensitivity warrant a resurvey in the following cycle.

Table 6.5: Was the sampling design adequate to determine the spatial extent of the effect?

If a relationship between the benthic community and the effluent has been detected, but the spatial extent of the relationship is not definable from the data collected, increased sampling effort at intermediate levels of exposure is required. For example, if only a near-field site and a far-far field site had been sampled and compared in a coastal mill study, and found to differ with respect to several community descriptors, the study should be augmented to a gradient or radial gradient design to better define the extent of the effluent effects. If only an upstream site and near-field site had been sampled and compared in a study of a mill located on a river, and the sites were found to differ with respect to several community descriptors, the study should be augmented to include more sites at varying distances downstream of the mill to better define the extent of the effluent effects. When the design is augmented as described, adequate sampling effort at each level of effluent exposure is required to precisely estimate the spatial extent of the effect.

Table 6.6: Alternative Approaches for Assessment of Benthic Community Impacts

Currently, there are no accepted alternative approaches to in situ (field) benthic surveys that can correct for confounding factors. However, there are several areas worth investigation as alternative approaches.

- Controlled field mesocosm bioassays are undergoing pilot tests for potential future application in estuarine environments. Examples of the use of mesocosms to assess pulp and paper mill effluents in freshwater environments include Hall and Haley (1991), Podemski and Culp (1996), Borton *et al.* (1996), Culp *et al.* (1996), and Saski *et al.* (1996).
- Artificial substrates may be appropriate in situations where suitable substrates cannot be found or sampled (e.g., historical contamination of substrates, inaccessible substrates, etc.), although this sampling method is not normally desirable.
- The use of the sediment quality triad (Chapman, 1990) could resolve confounding factors associated with both historical discharges and multiple discharges.
- Caged bivalve studies are not appropriate for the assessment of benthic community responses although they may be valid alternatives for species-specific responses as outlined for the Fish Survey EEM requirement.
- The RCA approach, when practical and appropriate, addresses much of the concern regarding appropriate reference sites and resolving the affects of confounding factors.

There is a need for industry and government to participate in the development and acceptance of alternative approaches.

7.0 Future work and Research Needs

7.1 Technical Guidance Document

The current Technical Guidance Document (TGD) is not in accordance with the recommendations from Cycle 1 proposed in this document. It is recommended that the TGD be revised by the BCEWG to assure that it is consistent with the recommendations of this report and the proposed decision tree tables.

7.2 Research and future work areas

As indicated in Section 5.0 there are no currently approved alternatives to the study designs being used for the benthic community surveys. However, some alternatives have been discussed by the BCEWG as having potential and being worthy of further study. These are: 1) the use of mesocosms; 2) the application of the reference condition approach (RCA); 3) the use of the sediment quality triad; 4) microbial characterization based on sediment chemistry (Eh and total sulphide) which may be related to invertebrate community responses; and 5) the examination of Cycle 1 data, where suitable, to determine if higher taxonomic levels can provide cost savings without compromising the sensitivity of the EEM assessment.

7.2.1 Mesocosm approach

Field biomonitoring in marine and freshwater environments is often hindered by confounding environmental factors such as the difficulties in estimating exposure (duration and concentration) to pollutants or the problem of multiple discharges simultaneously impacting the receiving waters. In these situations, inferential hypothesis testing based upon biomonitoring data can often be equivocal because of non-independence of treatments (Stewart-Oaten *et al.* 1992; Cooper and Barmuta 1993).

In contrast to field biomonitoring, experimentation using mesocosms provides replication as well as precise control over relevant variables, leading to uncompromised hypothesis testing and important insight into mechanisms of biotic responses to pollutants. Mesocosm approaches can offer a higher level of ecological complexity than acute toxicity tests (Forbes and Forbes 1994), thereby providing important ecological information concerning the chronic effects of pollutants on benthic food chains and food webs. The application of mesocosms essentially allows ecosystem-level questions to be posed and chronic-exposure treatments to be carried out at a smaller, more manageable scale than that of whole-ecosystem manipulations (Kimball and Levin 1985; Shaw and Kennedy 1996). Mesocosms range widely in design, from simple laboratory systems to elaborate outdoor complexes, and have been used to examine many levels of ecological organisation, including single species, multi-species assemblages and communities. Several recent articles have reviewed the use and application of mesocosms in ecotoxicology for riverine (Guckert, 1993; Pontasch, 1995; Culp *et al.*, 1996) or estuarine (Lauth *et al.* 1996) habitats.

Recent examples of the use of mesocosms to assess the effects of pulp and paper mill effluents stem from recent work for the Northern Rivers Basins Study (Podemski and Culp, 1996) and Fraser River Action Plan (Lowell *et al.*, 1995; Culp *et al.*, 1996). The development of a field-based stream mesocosm system for these large northern rivers provided the means by which a mechanistic understanding of the effects of BKME-related nutrient and contaminant stressors on the benthic biota could be obtained. Causality could be assigned definitively in a field application where inferential hypothesis testing was very limited. For example, the dominant effect of BKME discharge to the upper Athabasca and Fraser rivers was that of nutrient enrichment and stimulation of food-web productivity. These studies were restricted to investigating the effect of BKME on the benthic food web (i.e., periphyton to invertebrates), but other response variables such as fish growth or endocrine induction could also be examined using this approach. Mesocosm research is a promising technique for consideration in aquatic environmental effects monitoring programs for the pulp and paper industry because cause-and-effect scenarios can be investigated, and ecological indicators for riverine biota developed under experimentally controlled dose-response regimes. They would also be valuable tools for assessing the potential for additional mills along the river to raise overall contaminant and nutrient concentrations to levels that could degrade the ecosystem.

It is recommended that pilot studies be undertaken to evaluate the applicability to Cycle 2 studies of mesocosm approaches as an alternative method in situations where confounding factors prevent the use of approved field sampling designs.

7.2.2 Alternate sampling program designs using predictive modelling techniques: the reference condition approach (RCA).

Most study designs use one or more control sites sampled contemporaneously with potentially disturbed sites to test whether or not there have been changes in community structure. This study design philosophy assumes that the control sites are equivalent to the test sites in all respects except the stressor of concern. An alternate approach is to use a set of pre-sampled reference sites that capture the range of benthic communities that occur in a region. The test site is then compared to a subset of appropriate reference sites, representing the type of benthic community expected to be found at the test site. The selection of appropriate test sites can be made using a predictive model. The predictive model uses habitat attributes at the test site to predict which of the potential reference sites best match the test site. Comparison of the benthic community at the test site with the community found at the reference sites forms the basis for deciding the degree of impairment occurring at the test site.

For this approach to be valid for EEM four requirements must be met:

1. benthic communities integrate the environmental conditions at a site through time;
2. benthic communities respond to the stress of pulp and paper effluents;
3. variation in communities from site to site is predictable; and
4. reference and exposure sites, or degree of exposure at sites, can be defined.

Collection of a reference database is a significant undertaking. In the Fraser River watershed (234,000 km^2) 250 reference sites were required to capture the biological variability of the region; a similar number were required in the Great Lakes. Once the database is developed it can be used in perpetuity and for multiple uses; although this activity represents a considerable initial investment, it has long-term use. Such databases have multiple potential users (government, industry and consultants), can be used for both site-specific and regional assessment and planning, and so they should be developed and maintained by government agencies.

Development of regional databases nationwide must be a gradual process, so this approach cannot be implemented immediately in Cycle 2. Furthermore, it will be necessary to demonstrate that the RCA is a practical methodology. Accordingly, we see a need to improve the currently used control-impact approach for Cycle 2 but to modify it so that the data collected can be used in the RCA as the databases are developed.

It is recommended that the BCEWG investigate the usefulness of the RCA from Cycle 1 data for those mills where reference data bases exist and provide advice on their applicability to Cycle 2 studies.

7.2.3 Sediment Triad

The problem of confounding factors has been a recurring theme through the interpretation of Cycle 1 data. One approach to resolution of confounding factors associated with both historical discharges and multiple discharges could be the use of the sediment quality triad (Chapman, 1990). This assessment method incorporates chemical data, laboratory bioassays and community data in a decision-making framework that can distinguish historical from current disturbance and concerns associated with media-related stressors.

It is recommended that the usefulness of this method be examined in a pilot study.

7.2.4 Microbial Characterization

It is recommended that the microbial characterization method be further tested and linked to benthic macrofaunal response. Methods have been developed to predict the type of microbial and macrofaunal communities found in marine ecosystems in response to organic enrichment. The applicability of this approach to detecting the response of macroinvertebrate communities to pulp and paper mill discharges should be examined.

7.2.5 Taxonomic Resolution

It is recommended that Cycle 1 and Cycle 2 data be examined to determine whether higher- (e.g., family-) level identifications are sufficient to determine effects from mill discharges. If this is the case, it is recommended that only family-level identification be required in future rather than the more costly genus-level identifications.

8.0 References

Bailey, R.C., Day, K.E., Norris, R.H. and T.B. Reynoldson. 1995. Macroinvertebrate community structure and sediment bioassay results from nearshore areas of North American Great lakes. *J.Great Lakes Res.*, 21:42-52.

Borton, D.L, Streblow, W.R., Bradley, W.K., Bousquet, T., Van Veld, P.A., Wolke, R.E., and A.H. Walsh. 1996. Survival, growth, production and biomarker responses of fish exposed to high-substitution bleached kraft mill effluent in experimental streams. In *Environmental Fate and Effects of Pulp Mill Effluents*. M. Servos, J. Carey, K. Munkittrick, and G. Van Der Kraak (Eds.). St. Lucie Publisher, FL, pp. 473-481.

Chapman, P.M. 1990. The sediment quality triad approach to determining pollution-induced degradation. *Sci.Total.Envir.*, 97/98:815-825.

Clarke, K.R. 1993. Non-parametric multivariate analyses of changes in community structure. *Aust.J.Ecol.*, 18:117-143.

Clarke, K.R. and M. Ainsworth. 1993. A method of linking multivariate community structure to environmental variables. *Mar.Ecol.Prog.Ser.*, 92:205-219.

Cooper, S.D. and L. Barmuta. 1993. Field experiments in biomonitoring. In *Biomonitoring and Benthic Invertebrates*. D.M. Rosenberg and V.H. Resh (Eds.), Chapman and Hall, New York, pp. 399-441.

Corkum, L.D. and D.C. Currie. 1987. Distributional patterns of immature Simuliidae (Diptera) in northwestern North America. *Freshwat.Biol.*, 17:201-221.

Corkum, L.D. 1991. Spatial patterns of macroinvertebrate distributions along rivers in eastern deciduous forest and grassland biomes. *J.N.Am.Benthol.Soc.*, 10:358-371.

Cuffney, T.F., M.E. Gurtz, and M.R. Meador. 1993. Guidelines for the processing and quality assurance of benthic invertebrate samples collected as part of the national water-quality assessment program. U.S. Geological Survey Open-File Report 93-407.

Culp, J.M., Podemski, C.L., Cash, K.J., and R.B. Lowell. 1996. Utility of field-based artificial streams for assessing effluent effects on riverine ecosystems. *J.Aquat.Eco.Health*, 5: 117-124.

Environment Canada and Department of Fisheries and Oceans. 1992. Aquatic Environmental Effects Monitoring Requirements. Annex 1 to EPS 1/RM/18. May 20, 1992. 23 (+iii)pp.

Forbes, V.E. and T.L. Forbes. 1994. *Ecotoxicology in Theory and Practice*. Chapman and Hall, New York, 247pp.

Green, R.H. 1979. *Sampling Design and Statistical Methods for Environmental Biologists*. Wiley, NY, 257pp.

Guckert, J.B. 1993. Artificial streams in ecotoxicology. *J.N.Am.Benthol.Soc.*, 12:313-384.

Hall, T.J. and R.K. Haley. 1991. Effects of biologically treated bleached kraft mill effluent on cold water stream productivity in experimental streams. *Environ.Toxicol.Chem.*, 10:1051-1060.

Hamel, P., et G. Vaillancourt. 1997. L'utilisation du benthos à titre de biomoniteur dans les études du suivi des effets sur l'environnment (ESSEE). Présenté à l'association des industries forestières du Québec. January, 1997.

Hargrave, B.T., Duplisea, D.E., Pfieffer, E. and D.J. Wildish. 1993. Seasonal changes in benthic fluxes of dissolved oxygen and ammonia associated with marine cultured Atlantic salmon. *Mar.Ecol.Prog.Ser.* 96: 249-257.

Hargrave, B.T., Phillips, G.A., Doucette, L.I., White, M.J., Milligan, T.G., Wildish, D.J., and R.E. Cranston. Assessing benthic impacts of organic enrichment from marine aquaculture. *Wat.Air Soil Pollut.* (In press).

Holme, N.A. and A.D. McIntyre (Eds). 1984. *Methods for the Study of Marine Benthos,* Second Ed., Blackwell, Oxford, 387pp.

Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* 54: 187-211.

Johnson, R.K., Wiederholm, T., and L. Eriksson. 1991. The influence of season on the classification and ordination of profundal communities of nutrient poor, oligo-mesohumic Swedish lakes using environmental data. *Verh.Internat.Verein.Limnol.*

Kimball, K.D. and S.A. Levin. 1985. Limitations of laboratory bioassays: the need for ecosystem-level testing. *BioScience*, 35:165-171.

Lauth, J.R., Scott, G.I., Cherry, D.S. and A.L. Buikema, Jr. 1996. A modular estuarine mesocosm. *Environ.Toxicol.Chem.*, 15: 630-637.

Lowell, R.B., Culp, J.M. and F.J. Wrona. 1995. Stimulation of increased short-term growth and development of mayflies by pulp mill effluent. *Environ.Toxicol.Chem.*, 14: 1529-1541.

Lowell, R.B. 1997. Discussion paper on critical effect size guidelines for EEM using benthic invertebrate communities. *Report to Environmental Effects Monitoring Program*, Environment Canada, Ottawa, ON.

Ormerod, S.J. and R.W. Edwards. 1987. The ordination and classification of macroinvertebrate assemblages in the catchment of the River Wye in relation to environmental factors. *Freshwat.Biol.*, 17:533-546.

Parker, R. and N. Smith. 1997. A synopsis of the first cycle of the pulp and paper mill environmental effects monitoring program in the Atlantic region. February, 1997.

Podemski, C.L. and J.M. Culp. 1996. Nutrient and contaminant effects of bleached kraft mill effluent on benthic algae and insects of the Athabasca River, AB, Canada. In *Environmental Fate and Effects of Pulp Mill Effluents*. M. Servos, J. Carey, K. Munkittrick, and G. Van Der Kraak (Eds.), St. Lucie Publisher, FL, pp. 571-580.

Pontasch, K.W. 1995. The use of stream microcosms in multispecies testing. In *Ecological Toxicity Testing: Scale, Complexity and Relevance*. J. Cairns and B.R. Niederlehner (Eds.), Lewis Publishers, Boca Raton, FL, pp. 169-191.

Reynoldson, T.B., Day, K.E., Bailey, R.C., and R.H. Norris. 1995. Methods for establishing biologically based sediment guidelines for freshwater quality management using benthic assessment of sediment. *Aust.J.Ecol.*, 20:198-219.

Saski, E.K., Vahatalo, A. , Salonen, K., and M.S. Salkinoja-Salonen. 1996. Mesocosm simulation on sediment formation induced by biologically treated bleached kraft pulp mill wastewater in freshwater recipients. In *Environmental Fate and Effects*

of Pulp Mill Effluents. M. Servos, J. Carey, K. Munkittrick, and G. Van Der Kraak (Eds.), St. Lucie Publisher, FL, pp. 261-270.

Shaw, J.L. and J.H. Kennedy. 1996. The use of aquatic field mesocosm studies in risk assessment. *Environ.Toxicol.Chem.*, 15: 605-607.

Stewart-Oaten, A., Bence, J.R., and C.W. Osenberg. 1992. Assessing effects of unreplicated perturbations: no simple solutions. *Ecology*, 67:929-940.

Warwick, R.M. 1993. Environmental impact studies on marine communities: Pragmatical considerations. *Aust.J.Ecol.*, 18:63-80.

Warwick, R.M. and K.R. Clarke. 1993. Comparing the severity of disturbance: a meta-analysis of marine macrobenthic community data.. *Mar.Ecol.Prog.Ser.*, 92:221-231.

Wildish, D.J. 1983. Coastal zone management and the pulp and paper industry. *Pulp and Paper Canada*, 84:T145-T148.

Wright, J.F., Moss, D., Armitage, P.D., and M.T. Furse. 1984. A preliminary classification of running-water sites in Great Britain based on macroinvertebrate species and the prediction of community type using environmental data. *Freshwat.Biol.*, 14:221-256

Wrona, F.J., Culp, J.M., and R.W. Davies. 1982. Macroinvertebrate subsampling: a simplified apparatus and approach. *Can.J.Fish.Aquat.Sci.*, 39:1051-1054.